

Care Quality, Technology Use, and Online Patient Perception

Danish H. Saiffee
University of Alabama, Tuscaloosa
dhsaiffee@cba.ua.edu

Atanu Lahiri
University of Texas, Dallas
atanu.lahiri@utdallas.edu

Indranil R. Bardhan
University of Texas, Austin
indranil.bardhan@mcombs.utexas.edu

Zhiqiang (Eric) Zheng
University of Texas, Dallas
ericz@utdallas.edu

Abstract

The literature on the efficacy of online reviews suggests that such reviews are usually effective in informing consumers about the product or service. This mitigates information asymmetry, paving the way for an efficient marketplace. However, the literature is unclear about the usefulness of online reviews in the healthcare context. Since healthcare is largely a credence good, it is indeed possible that online reviews are not as informative in its case as they are in some others. In this work, we take a closer look at what online physician reviews actually capture, by studying the association between online reputation of a physician and her adherence to clinical guidelines. We also study the association between reputation and electronic health record (EHR) usage. Our results reveal that online reputation does not adequately reflect care quality, in the sense that improved adherence to care guidelines does not seem to be associated with better online reviews. However, EHR usage seems to have a somewhat positive association, suggesting that reviews can capture efficiency improvements from information technology even when they do not capture care quality.

Keywords: Online reviews, clinical guidelines, adherence, physician quality reporting, electronic health records, text mining, online reputation, care quality, efficiency.

1. Introduction

There is no clear consensus among researchers about the efficacy of online reviews in the context of healthcare. On one hand, Gray et al. [13] did not find evidence of any association between physician website ratings and clinical outcomes, such as blood pressure or low-density lipoprotein controlled. This lack of association is also echoed in the work by

Saiffee, et al. [25]. On the other, Lu and Rui [18], who examined the validity of online (star) ratings of cardiac surgeons in the context of coronary artery bypass graft (CABG) surgeries, found that patients operated by cardiac surgeons with a lower rating do exhibit a higher likelihood of mortality vis-à-vis those with a higher rating or no rating. Likewise, Bardach et al. [1] also documented a favorable association between patient ratings on Yelp and clinical outcomes, such as the readmission and mortality rate.

Notwithstanding questions over their reliability, online reviews of physicians continue to rise in popularity. According to a recent study [15], approximately sixty percent of patients now consider online reviews to be an important factor in their selection of physicians. Moreover, in a recent survey [2], twenty-eight percent strongly agreed that a positive online review of a physician would cause them to seek care from that physician, while another twenty-seven percent indicated that a negative review would cause them to avoid that physician altogether. Interestingly, many physicians have also begun to monitor their online reviews closely, looking for ways to boost their ratings on review sites such as Yelp, Vitals, and RateMDs. There have even been instances in which physicians have filed defamation lawsuits over negative patient reviews [25].

In this backdrop, we take a closer look at online reviews, and ask what they actually capture. Do they capture care quality, accurately reflecting the clinical performance of the physician? Further, even if unable to do so, are they at least capable of capturing efficiency improvements brought about by increased technology usage? Note that these two questions are clearly different, which makes both quite important. It has been argued that healthcare is a credence good [11]. Although the quality of a credence good is difficult to assess—for example, there is no easy way for many of us to figure out whether a repair made by an auto-repair shop was actually necessary or done the right way—we may still be able to observe the wait

times, attentiveness of the staff, and other elements indicative of a very efficient provider. In this work, we empirically examine if this could also be the case with healthcare, by drawing on several rich data sources on physician quality reporting as well as patients' online reviews of their physicians.

To measure online reputation, we use textual reviews as well various numeric/star ratings available at Vitals.com. Online reputation, our main dependent variable, can then be measured by either the sentiments expressed in these reviews or their numeric ratings. To obtain data on care quality and EHR usage, we leverage the Physician Quality Reporting System (PQRS).

The PQRS was started by the Centers for Medicare and Medicaid Services (CMS) to streamline numerous quality-reporting and measurement programs at hospitals and group practices across the country. The PQRS collects data on various dimensions of care delivery, including prevention, care coordination, and resource utilization. This program uses financial incentives to encourage eligible professionals/providers (EPs) to report data on the quality of healthcare delivery. When the PQRS was originally implemented in 2006, it financially rewarded providers who volunteered to disclose quality information. In 2015, however, the program has moved from a voluntary program to a mandatory one, imposing penalties on providers for failing to participate. Starting in 2017, the PQRS is no longer a stand-alone program, and has become an integral part of the Merit-based Incentive Payment System (MIPS). Nevertheless, the penalty remains in force—a failure to satisfactorily report data on healthcare quality measures results in negative payment adjustments via the PQRS. For physicians who participate, the CMS uses the reported information to measure their adherence to clinical guidelines (or “performance rate” as it is commonly called) as relevant to their respective specialty.

To answer our broader research question, we investigate the following: (1) Do physicians who adhere more closely to evidence-based clinical guidelines receive better online reviews? (2) Do physicians who use EHR systems receive better online reviews? (3) Does the use of EHR by a physician positively moderate the relationship between her adherence to clinical guidelines and the online reviews that she receives? The first question is about the relationship between care quality and online reputation, while the second one is about the association between technology usage and online reputation. The third one is to consider the fact that a potential moderator of the relationship between physicians' adherence to clinical guidelines and their

online reviews could be their usage of EHR. This is because EHR can improve the flow of information within the provider organization, as well as across organizations and patients, making it easier to meet the CMS-specified guidelines.

To recognize the relevance of EHR in our context, note that, as of the end of 2016, over sixty percent of all US office-based physicians have demonstrated meaningful use of certified health information technology (IT) in the CMS-sponsored EHR incentive programs. Adoption and usage of EHRs by physicians have the potential to alter the workflow of their daily work and their interaction with patients, which might, in turn, impact their relationship with their patients. In light of the increasing role of EHRs in physicians' daily routine and work, it becomes exceedingly important to take into account both the direct and moderating roles of this technology, and we certainly do so here. To the best of our knowledge, this is one of the first studies to empirically address all these issues—care quality, efficiency from technology usage, and online reputation—at the physician-level through a unified framework.

We find that there is no direct relationship between physician adherence to clinical guidelines and the online reviews. We also find that physicians who use EHRs tend to receive more positive textual feedback as well as somewhat higher online ratings. We did not find any moderating effect of the use of EHRs by physicians on the relationship between their adherence to clinical guidelines and online reviews. The lack of connection between physician adherence to clinical guidelines and online reviews is consistent with the credence nature of healthcare services. This aspect of healthcare services can make it quite difficult for patients—who lack the deep and broad knowledge about medical care provided by their physicians—to effectively evaluate the quality of care provided. In contrast, efficiency gains from EHR are clearly more tangible to them, and have very clear ramifications on the quality and depth of reviews that they write.

2. Literature Review

Our study lies at the intersection of: (1) the broader research on consumer reviews, (2) the more closely-related stream of online physician reviews, (3) the literature on adoption and use of health IT, particularly EHRs, and finally, (4) the economics literature on credence goods.

2.1 Consumer Reviews

Over the past two decades, the first stream—the literature on consumer reviews in general—has grown

considerably. One branch of this research has examined supply or production of consumer reviews and addressed a wide variety of questions, including word-of-mouth and social influence [7, 22], evolution of ratings [12], reporting biases [11, 17], and review manipulation [19].

The demand or consumption side has also received considerable attention. It has been found that not only do potential consumers read and rely on such reviews, but the product ratings in many cases also correlate with the actual sales [3, 4, 5, 9, 14]. Finally, researchers have also focused on the issue of perceived and actual usefulness of online reviews [10, 18, 21].

Although we do not contribute to the branch on consumption of reviews, we contribute to the other two. Specifically, we add to the literature on production of reviews by examining whether the review production process faithfully captures what physicians actually do. Also, by examining the efficacy of reviews in the healthcare context, we are able to extend the broader research on usefulness of online reviews.

2.2 Online Physician Reviews

Closely related to this study are the papers that examine the association between objective measures of care quality and online reviews [1, 13, 18, 25]. As mentioned already, there is no agreement among researchers in this stream about the usefulness of online reviews, and further research is clearly necessary.

Notwithstanding the papers above, there is a scarcity of research on the managerial, economic, and behavioral antecedents of online healthcare reviews, with respect to the question of what reviews actually capture. We show that online reviews can capture gains from technology usage even when they are unable to capture the care aspects. This way, we advance the research on physician reviews. Besides, there are very few studies that have looked at all three aspects—clinical performance, EHR usage, and patient perception—at the physician level. Such physician-level analyses are highly relevant because insights gained from hospital-level studies may not apply to individual physicians [25].

Moving on to research on the PQRS, the PQRS program offers providers an opportunity to assess and compare their performance vis-à-vis their peers. However, there happens to be very limited literature on these aspects of the PQRS. It is also unclear whether physicians see enough merit in this program. For instance, a survey of physicians by Federman and Keyhani [8] found that a majority of participating physicians were skeptical about the impact of the

PQRS on the quality of care provided by them to their patients. In other words, the value of the PQRS to physicians is still very much an open question, notwithstanding the faith shown by the CMS in the merits of the PQRS. Our work is one of the first to study whether adherences to clinical guidelines stipulated by the PQRS potentially translates to a better patient perception.

2.3 Use of EHRs by Physicians

Even though many studies have been conducted over the past two decades to examine various aspects of adoption and use of health IT (HIT) systems at the hospital level, there is still a lack of clear understanding on their benefits to physicians. This is partly because it has been difficult to obtain HIT usage data at the physician level. Fortunately, however, there has been a significant push towards rapid adoption and use of HIT since the enactment of the HITECH Act. This push has improved the access to data on the use of EHRs, making this study possible.

EHRs can not only help physicians but also assist staff members in building a stronger professional relationship with patients. For instance, a review study [16], which is based on prior studies on the use of EHRs by physicians in outpatient settings, finds evidence of a positive or neutral relationship between EHR usage and patient satisfaction, in six of the seven articles, and a negative relationship in one study. In our research, we examine whether there is any relationship between EHR use by physicians and their online reviews. In other words, we extend the existing literature on EHR use by physicians and their relationship with their patients to the online review setting, a setting that is continually becoming more relevant to the healthcare sector.

2.4 Credence Goods

The term credence was originally suggested by Darby and Karni [6] to characterize goods and services whose quality information is never revealed to consumers. Since healthcare happens to be a very complex combination of multiple clinical steps—including, but not limited to, accurate and timely diagnosis, educating patients about relevant medical conditions, carrying out required medical procedures, following up with patients, and developing a comprehensive care plan—assessment of the quality of care is often quite challenging from the patient's perspective.

More recently, it has been suggested that consumer reviews and word-of-mouth could play a significant role in the context of professional services

that possess significant credence qualities [11]. Whether online ratings actually contain information on credence qualities is, however, an open question. Our work complements existing studies in this domain [e.g., 13, 18] by investigating if, indeed, online reviews are associated with physicians' adherence rate to clinical guidelines. We also shed some light on whether the use of EHRs by physicians improves their patients' experience.

3. Research Framework

Figure 1 displays our conceptual model which comprises three research questions. The first question, RQ1, is to understand whether online reviews reflect *quality* or *effectiveness* of care, as measured by the adherence rate to the clinical guidelines stipulated in the PQRS program. Since the PQRS guidelines were designed by the CMS after extensive consultations with experts with the objective of improving the quality of care, the adherence rate is a reasonable proxy for our purpose. In fact, our use of this proxy complements prior research that has used outcome-based proxies [e.g., 28].

The second question, RQ2, is to see whether online reviews reflect some of the *efficiency* gains that

we expect from EHR usage. The data dictionary provided by the CMS does suggest that the use of EHR helps clinicians provide efficient and beneficial care, by making information available in a timely manner. So, the association of efficiency gains from EHR use with online reviews ought to be of particular interest.

The motivation for the third question, RQ3, is as follows. EHR systems can help physicians follow clinical guidelines more closely, which in turn, might strengthen the relationship between clinicians and their patients. For example, EHR systems can remind a physician and her staff to follow up with patients regularly, remind physicians about recommended tests, help the clinicians adhere to the care plan for their patients, and monitor treatment. Hence, it is possible that EHR use by physicians can positively moderate the relationship between clinical adherence rate and their online reputation. Whether or not this moderation is significant is precisely what RQ3 is aimed at finding out.

A note is in order. Above, we are using the terms *effectiveness/quality* and *efficiency* in a rather limited sense. However, our usage is consistent with broader literature [e.g., 24] and appropriate in the healthcare context [e.g., 20].

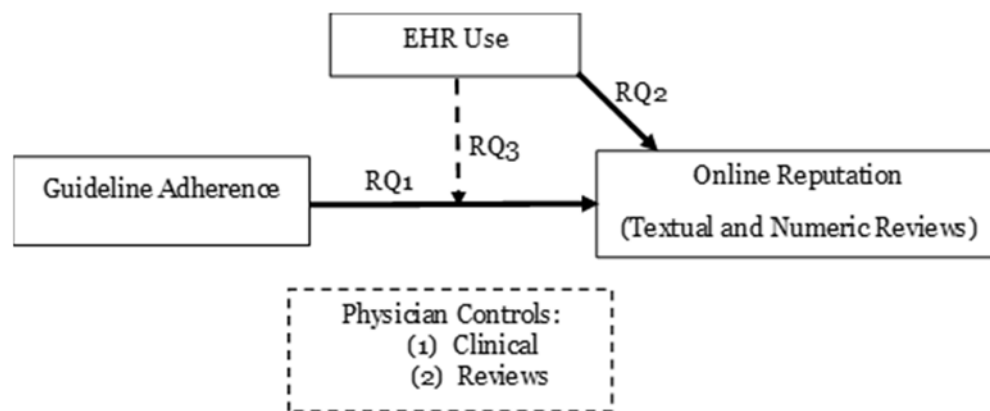


Figure 1. Conceptual Framework

4. Data and Variable Construction

We obtained research data from three sources: (1) Vitals.com, a public physician review platform, (2) Individual Eligible Provider Public Reporting and Physician Compare National Downloadable File (2015), and (3) Provider Utilization and Payment Data (Physician and Other Supplier Public Use File). The first dataset provides us data on online reviews of physicians, including textual comments and star ratings provided by patients. We use this dataset to

construct our dependent variables. The dataset contains physicians from North Texas, and spans a nine-year period from 2007 to 2015. A key aspect of our analysis is that we integrate this dataset with the other two, using physician attributes as matching identifiers.

The second dataset was collected from the CMS. Both of our key explanatory variables, *EHR_use* and *ClinicalAdherenceRate* (description given below), were constructed from this dataset. The third dataset was also collected from the CMS. It was prepared by

the CMS to share information on services and procedures provided to Medicare beneficiaries by various physicians and other healthcare professionals. It contains information on utilization, payment (allowed amount and Medicare payment), and submitted charges, organized by National Provider Identifier (NPI), Healthcare Common Procedure Coding System (HCPCS) code, and the location of service. The details are based upon the CMS administrative claims data for Medicare beneficiaries enrolled in the program. This dataset also includes provider demographics (name, credentials, gender, complete address, and entity type) from the National Plan and Provider Enumeration System (NPPES).

4.1 Dependent Variables

The main dependent variables are *SentimentScore*, *OverallRating*, *PhysicianRating* and *StaffRating*. These are constructed based on the data collected from Vitals.com, where a patient assigns a star rating, on a 5-point scale, to his physician for several categories, in addition to writing a textual review. *OverallRating* is the average of the star ratings received by a physician, and *SentimentScore* is the average of the sentiment scores derived from textual reviews.

To construct the variable, *OverallRating*, we proceed as follows: We first averaged all patient ratings corresponding to a review, across all categories: ease of scheduling appointments, promptness, friendliness of staff, accurate diagnosis, bedside manners, the time spent with patients, and appropriate follow-ups. The *OverallRating* of a physician is then the average of this category average across all her reviews. The construction of the variables *PhysicianRating* and *StaffRating* is similar to that of *OverallRating*, except that we use only four of the seven categories (accurate diagnosis, bedside manners, the time spent with patients, and appropriate follow-ups) for *PhysicianRating* while we use the remaining three (ease of scheduling appointments, promptness and friendliness of staff) for *StaffRating*.

To construct *SentimentScore*, we employed sentiment analysis. This technique classifies the valence of each sentiment word in a review into four sentiment categories: very positive, positive, negative, and very negative, based on the vocabulary provided by Nielsen [23]. Then, an aggregation across all sentiment words within a review yields the sentiment score of the review; while aggregating, we treat very positive as +2, positive as +1, negative as -1, and very negative as -2. The *SentimentScore* for a physician is then the average of the sentiment score across all her reviews.

Besides the aforementioned variables, we also use the individual components of *PhysicianRating* and *StaffRating* as dependent variables. These seven dependent variables are: *DiagnosisRating* (accurate diagnosis by physician), *FollowupRating* (appropriate/timely follow-ups by physician), *SpendsTimeRating* (time spent by the physician with a patient), *BedsideMannersRating* (bedside manners of the physician), *StaffCourtesyRating* (friendliness of the staff), *PromptnessRating* (staff promptness) and *AppointmentEaseRating* (ease of scheduling appointments for a patient).

4.2 Key Explanatory Variables

EHR_use takes a value of one for physicians who successfully participate in the EHR Incentive Program, and zero for others. *ClinicalAdherenceRate* is the average of the performance scores (out of 100) for the clinical guideline measures reported by a given physician in the PQRS. Note that these performance scores are assigned by the CMS to clinicians, and that they are based on the information reported by eligible physicians participating in the PQRS.

4.3 Control variables

Clinical Controls: For each physician, *ClinicalAdherenceMeasureCount* is the number of clinical guideline measures that the physician reports to the CMS. *NumPatientsClinicAdherenceMeasure* is the size of the patient population that forms the basis for the numbers reported by the physician. *GraduationYear* indicates the year of her graduation from medical school. *GenderFemale* equals zero if the physician is male, and one otherwise. *NumHCPCS* is the total number of unique HCPCS (Healthcare Common Procedure Coding System) codes. *NumServices* denotes the number of services provided by her. *NumMedicareBeneficiaries* is the number of Medicare beneficiaries receiving her services. *BeneficiaryAvgAge* is the average age of the beneficiaries, as determined at the end of the calendar year or at the time of death. *BeneficiaryAvgRiskScore* is the average hierarchical condition category (HCC) risk score of the beneficiaries.

Review Controls: *ReviewWordsNum* represents the average length of all reviews available for the physician. This explanatory variable allows us to differentiate between reviews that might have the same *SentimentScore* but differ in terms of their depth [21]. For instance, a review with ten positive terms, six negative terms, and a very negative term will have the same *SentimentScore* as a review with only two positive terms, but the former review is surely more

detailed in terms of sentiment content than the latter is. Further, to incorporate the possibility that textual content with high levels of negative word-of-mouth may coexist with high levels of positive word-of-mouth, we construct *SentimentVariance*. This variable is the average of the sentiment variance within a review, average taken across all reviews of the physician. *NumberOfTextualComments* controls for the number of textual reviews received by a physician.

The textual content in each review can be summarized by one or multiple latent topics/themes. To identify the topics, we conduct fine-grained textual analyses of the online reviews using a Latent Dirichlet Allocation (LDA) method. We construct four latent topics to represent the most dominant theme in each review, as shown in Table 1, based on which we then construct the following control variables. For a given physician, *TopicSurgery* represents the proportion of reviews for which the most prominent latent theme is the surgical and procedural competence of her service. *TopicOverallCare* is the proportion of her reviews where the most prominent underlying theme is the overall care provided by her and her staff. *TopicPromptness* is the proportion of reviews where the most dominant theme is promptness of the staff in

the physician's office (wait times, appointment-scheduling experience, etc.). The baseline variable against which these three variables are interpreted is *TopicPhysician*. *TopicPhysician* is essentially the proportion of reviews where the most prominent theme is the physician herself (which includes time spent by her, as well as the treatment offered and her listening skills).

Table 1. Latent Topics Based on Closest Stems

<i>Surgery</i>	<i>Physician</i>	<i>Promptness</i>	<i>OverallCare</i>
surgeri	doctor	offic	staff
pain	patient	call	care
life	time	wait	recommend
back	year	appoint	great
year	good	time	feel
procedur	care	nurs	always
sever	medic	back	question
work	problem	anoth	friend
result	ever	hour	high
right	know	rude	concern

Table 2 provides the essential summary statistics of all our variables.

Table 2. Descriptive Statistics

Variable Type	Variable	Mean	Median	Std. Dev.	Min	Max
Aggregated Dependent Variables	<i>SentimentScore</i>	1.816	2.000	2.513	-8	12
	<i>OverallRating</i>	3.905	4.143	1.101	1	5
	<i>PhysicianRating</i>	3.893	4.200	1.193	1	5
	<i>StaffRating</i>	3.972	4.167	1.029	1	5
Individual Physician-level Dependent Variables	<i>DiagnosisRating</i>	4.020	4.429	1.181	1	5
	<i>FollowupRating</i>	3.729	4.000	1.323	1	5
	<i>SpendsTimeRating</i>	3.914	4.297	1.240	1	5
	<i>BedsideMannersRating</i>	3.944	4.372	1.230	1	5
Individual Physician-level Staff Variables	<i>StaffCourtesyRating</i>	4.010	4.326	1.123	1	5
	<i>PromptnessRating</i>	3.856	4.000	1.123	1	5
	<i>AppointmentEaseRating</i>	4.079	4.375	1.046	1	5
Key Explanatory Variables	<i>EHR_use</i>	0.498	0.000	0.500	0	1
	<i>ClinicalAdherenceRate</i>	82.566	91.917	21.819	0	100
Clinical Controls	<i>ClinicalAdherenceMeasureCount</i>	2.814	2.000	1.953	1	14
	<i>NumPatientsClinicAdherenceMeasure</i>	273.660	144.500	363.268	20	4043
	<i>GraduationYear</i>	1990.283	1991.000	9.892	1958	2015
	<i>GenderFemale</i>	0.182	0.000	0.386	0	1
	<i>NumHCPCS</i>	58.894	50.000	38.885	1	302
	<i>NumServices</i>	5929.344	2600.000	14894.399	40	324291
	<i>NumMedicareBeneficiaries</i>	687.469	494.000	684.874	24	5931
	<i>BeneficiaryAvgAge</i>	72.165	73.000	3.316	53	83
Review Controls	<i>BeneficiaryAvgRiskScore</i>	1.596	1.421	0.665	0.588	6.004
	<i>SentimentVariation</i>	0.164	0.143	0.105	0	0.8
	<i>NumberOfWords</i>	57.961	50.333	46.558	1	722
	<i>TopicPhysician</i>	0.236	0.125	0.296	0	1
	<i>TopicSurgery</i>	0.245	0.125	0.306	0	1
	<i>TopicOverallCare</i>	0.261	0.167	0.318	0	1
	<i>TopicPromptness</i>	0.258	0.167	0.315	0	1
	<i>NumberOfTextualComments</i>	5.686	3.000	8.605	1	108

5. Estimation Model and Strategy

We use the following regression model to answer the questions RQ1, RQ2, and RQ3:

$$\begin{aligned} & \text{OnlineReputation}_i \\ &= \alpha \text{EHR_use}_i + \beta \text{ClinicalAdherenceRate}_i \\ &+ \gamma \text{EHR_use}_i \times \text{ClinicalAdherenceRate}_i \\ &+ \delta \text{Controls}_i + \varepsilon_i \end{aligned}$$

The dependent variable *OnlineReputation_i* represents either of the following for physician *i*: *SentimentScore*, *OverallRating*, *PhysicianRating*, *StaffRating*, or one of the seven components of the overall rating described in Section 3.1. *EHR_use_i* is the dummy that indicates whether physician *i* successfully participated in the EHR incentive program. *ClinicalAdherenceRate_i* is the average of the performance scores (out of 100) over all measures reported by physician *i*. The product of *EHR_use_i* and *ClinicalAdherenceRate_i* is essentially the interaction effect between *EHR_use_i* and *ClinicalAdherenceRate_i*, and it captures the possible moderating effect. To see the impact of this interaction term, we use an incremental approach: we first run our estimation without this term, and then we run the same again with this interaction term added. *Controls_i* represents the clinical and review controls for each physician, and ε_i is the idiosyncratic error.

One simple strategy of estimating the above equation would be to use the ordinary least squares (OLS) method. Although we have run this method, we do not consider it reliable. In particular, there may be some unobserved physician attributes driving both *ClinicalAdherenceRate_i* and *OnlineReputation_i*. To account for the possible bias due to such omitted variables, we construct an instrument variable (IV) named *PeerClinicalAdherenceRate_i*, the average rate of clinical adherence by the physician's peers. A physician's peer is a physician who provides service in the same hospital where the focal physician practices. We construct *PeerClinicalAdherenceRate_i* by matching the associated hospitals using their CMS certification number. The rationale for this instrument variable is that the performance of a physician with respect to clinical adherence guidelines is likely to be associated with the performance of her peers, though peer adherence by itself is unlikely to determine her online reputation.

We also create an IV for the interaction term between *EHR_use_i* and *ClinicalAdherenceRate_i*. This IV is essentially the interaction between *EHR_use_i* and *PeerClinicalAdherenceRate_i*. We, however, do not construct an instrument for *EHR_use_i*, the rationale being that a physician cannot unilaterally adopt EHR

without proper coordination with the hospital that she is associated with. In other words, the variable *EHR_use* is reasonably exogenous. Accordingly, we employ the two-stage least squares (2SLS) method. This approach is significantly more reliable when compared to the straightforward OLS.

6. Results

From the first stage regression, we found that the aforementioned IVs are quite strong for their respective endogenous variables. We list below the results of the second stage of the 2SLS specifications in Table 3. For brevity, we have only included the coefficients and the corresponding errors of our key explanatory variables, and left out the controls. Model (a) shows the estimation results without the interaction term, and (b) shows that with the interaction term.

As evident from Table 3, there is no statistically significant relationship between *ClinicalAdherenceRate* and *SentimentScore*. This implies that, even if a physician adheres closely to her respective clinical guidelines, it is not likely that she will receive better textual comments. The results are similar for *OverallRating*, *PhysicianRating*, and *StaffRating* as well—they, too, do not reflect the care quality as measured by adherences to the PQRS guidelines. Finally, all seven components of the physician and staff ratings fail to exhibit any statistically significant connection with adherence, again echoing the fact that online reputation might not accurately reflect the quality of a credence good. Overall, the answer to RQ1 thus happens to be a resounding 'no.'

Next, we move on to our question RQ2, which concerns the efficiency gains from technology usage. Interestingly, *SentimentScore*, *FollowupRating*, *StaffCourtesyRating*, *SpendsTimeRating*, and *BedSideMannersRating* are all positively associated with *EHR_use*. Thus, it seems that the use of EHR is perhaps improving the communication between patients and the physician as well as the patient perception with respect to the physician's staff. This is interesting. Although a patient might not recognize that the physician is adhering to clinical guidelines, the patient might still be able to perceive some gains from technology use, at least to the extent EHR frees up the physician and her staff from routine paperwork and allows for more interaction with the patient, or allows a more effective follow-up.

Moving on to RQ3, as evident from Table 3, the interaction terms are insignificant across the board. There is still no evidence that reviews can capture improved clinical adherence, directly or otherwise.

Table 3. 2SLS Estimation (Second Stage) Results

	Without interaction (a)	With interaction (b)	Without interaction (a)	With interaction (b)
	<i>SentimentScore</i>		<i>OverallRating</i>	
<i>EHR_use</i>	0.389** (0.169)	0.412** (0.178)	0.087 (0.069)	0.088 (0.068)
<i>ClinicalAdherenceRate</i>	0.125 (0.586)	-0.385 (0.694)	0.048 (0.204)	0.019 (0.232)
<i>EHR_use * ClinicalAdherenceRate</i>		1.708 (1.274)		0.100 (0.454)
	<i>PhysicianRating</i>		<i>StaffRating</i>	
<i>EHR_use</i>	0.118 (0.076)	0.119 (0.076)	0.053 (0.065)	0.054 (0.068)
<i>ClinicalAdherenceRate</i>	0.103 (0.229)	0.046 (0.257)	-0.010 (0.217)	0.063 (0.241)
<i>EHR_use * ClinicalAdherenceRate</i>		0.197 (0.513)		-0.292 (0.529)
	<i>Individual Physician Dimensions</i>		<i>Individual Staff Dimensions</i>	
	<i>DiagnosisRating</i>		<i>StaffCourtesyRating</i>	
<i>EHR_use</i>	0.094 (0.078)	0.095 (0.078)	0.123* (0.072)	0.123* (0.073)
<i>ClinicalAdherenceRate</i>	0.198 (0.245)	0.155 (0.271)	-0.057 (0.236)	-0.037 (0.253)
<i>EHR_use * ClinicalAdherenceRate</i>		0.151 (0.558)		-0.083 (0.589)
	<i>FollowupRating</i>		<i>PromptnessRating</i>	
<i>EHR_use</i>	0.157* (0.087)	0.161* (0.086)	0.031 (0.075)	0.035 (0.085)
<i>ClinicalAdherenceRate</i>	-0.020 (0.248)	-0.116 (0.288)	-0.117 (0.252)	0.091 (0.267)
<i>EHR_use * ClinicalAdherenceRate</i>		0.323 (0.531)		-0.832 (0.744)
	<i>SpendsTimeRating</i>		<i>AppointmentEaseRating</i>	
<i>EHR_use</i>	0.161** (0.082)	0.163** (0.081)	-0.030 (0.070)	-0.030 (0.071)
<i>ClinicalAdherenceRate</i>	0.072 (0.260)	0.020 (0.289)	-0.020 (0.248)	-0.022 (0.285)
<i>EHR_use * ClinicalAdherenceRate</i>		0.180 (0.597)		0.010 (0.530)
	<i>BedsideMannersRating</i>			
<i>EHR_use</i>	0.171** (0.079)	0.174** (0.080)		
<i>ClinicalAdherenceRate</i>	0.121 (0.240)	0.016 (0.273)		
<i>EHR_use * ClinicalAdherenceRate</i>		0.357 (0.526)		

p<0.1*, p< 0.05**, p<0.01***, heteroscedasticity-consistent standard errors within parentheses

Finally, a point is in order about the observed associations or a lack thereof. The reported lack of association between adherence and various staff dimensions in Table 3 is not entirely unexpected, as it is unlikely that guideline adherence by a physician has any material effect the performance of her staff. However, the same is not true for the other rating dimensions. In particular, it is interesting that *DiagnosisRating* has no association with adherence. The lack of associations with *FollowupRating*, *SpendsTimeRating*, and *BedSideMannersRating* are also quite telling. For instance, in the PQRS guideline list, there are several clinical adherence measures that involve follow-up, for example, (i) Biopsy Follow-up, (ii) Body Mass Index (BMI) Screening and Follow-Up Plan, and (iii) Screening for High Blood Pressure and Follow-up Documented. Likewise, there are several measures that account for communications between a physician and her patients on different aspects of care. These measures account for the time spent by the

physician in educating her patients and the bedside conversations. Despite the PQRS guidelines being so closely tied to the care given by the physician, we fail to find any association between adherence and physician rating dimensions. This lack of association can thus be considered indicative of the credence nature of healthcare.

7. Conclusions

A phenomenon intimately related to rapid adoption of social media is the rising popularity of online platforms that can make detailed reviews of physicians available to healthcare consumers. However, what these platforms actually offer consumers has remained a mystery, and prior researchers have reached conflicting conclusions. In this study, we reexamine the antecedents of online physician reviews with a twin focus. First, we examine whether reviews can capture the details of care

delivery and provide an indication of the quality of care. We do so by examining the association between online reputation and adherence to clinical guidelines stipulated by the PQRS. Second, we investigate whether the reviews can reflect information other than care quality itself, such as the efficiency, promptness, and courteousness of the physician and her staff. For this, we study the relationship between the use of EHRs by physicians and their online reviews. This unified research framework makes a unique contribution to the current literature on health IT systems, physician performance, healthcare quality, and patient perception in online social media.

Our key finding is that there is no direct relationship between adherences to clinical guidelines by physicians and their online reviews. This is consistent with the credence nature of healthcare services—it can indeed be quite difficult for an average healthcare consumer to objectively evaluate the quality of various clinical processes and services. It is true that stricter adherence to clinical guidelines stipulated in the PQRS program is an important dimension of the quality of care delivered by the physicians, especially in the current value-based payment era. At the same time, however, it seems unlikely that patients are able to observe or comprehend the actions of physicians with respect to adherence, let alone evaluate the benefits accrued therefrom. This finding is in stark contrast to what has generally been found in the context of services and products other than healthcare (such as books, movies, or music) and their online reviews. Clearly, healthcare is different, and so is the usefulness of reviews in the context of healthcare.

We do find a positive relationship between the use of EHRs by physicians and certain dimensions of online reviews. This finding is a reminder that, while patients may be unable to evaluate a physician's adherence and the quality of care delivered, they may still be able to observe some efficiency improvements. For example, to the extent that EHR usage allows a physician to spend more time with her patients, and provides efficient appointment scheduling or ease of access to their health data, patients might reciprocate by writing better online reviews. Nevertheless, we do not find any moderating influence of EHRs on the relationship between clinical guideline adherence by physicians and their online reviews, further reinforcing the notion that the quality of care is not as easy to assess as are tangible workflow improvements.

One way to relate to our findings would be to consider the case of an auto repair shop. It may not be easy for a customer there to figure out whether the shop accurately described the problem with his car and provided the optimal fix, but the customer may still be

able to discern that the mechanic was courteous and the service was prompt. The overall implication is thus that online reviews are at best partly informative and that we cannot overly rely on them to ensure a smooth transition to patient-centered care and an efficient healthcare marketplace. Rather, policymakers must be proactive in sharing any information that they possess, for example, the PQRS data, to the general public to keep them better informed. Perhaps, policymakers can also make some visualization tools available, since raw data might not appeal to consumers who are so used to intuitive star ratings. If our study is any indication, additional care quality data will nicely supplement what healthcare consumers can glean from existing online reviews. Bridging the information gap, in fact, would be critical to ensuring transparency and preventing patients from choosing physicians based on incomplete information, and doing so might also bring some comfort to physicians who have grown wary over the proliferation of online review sites.

In future research, we intend to conduct additional analyses to account for possible manipulations of online reviews. Another limitation of our study is that even though participation in the PQRS program is now mandatory and carries a penalty for non-participation, a significant number of providers are not yet participating in the PQRS program. We need to account for the possible self-selection bias that may result from this behavior. Addressing these gaps will lead to robust results, providing consumers, policymakers, and providers a clearer picture of the true usefulness of online physician reviews.

References

- [1] Bardach, N.S., R. Asteria-Peñaloza, W.J. Boscardin, and R. A. Dudley, "The Relationship between Commercial Website Ratings and Traditional Hospital Performance Measures in the USA", *BMJ Quality & Safety*, 2012 (22:3), pp. 194-202.
- [2] Burkle, C.M., and M.T. Keegan, "Popularity of Internet Physician Rating Sites and their Apparent Influence on Patients' Choices of Physicians", *BMC Health Services Research*, 2015 (15:1), pp. 416-422.
- [3] Chen, Y., and J. Xie. "Online Consumer Review: Word-of-mouth as a New Element of Marketing Communication Mix", *Management Science*, 2008 (54:3), pp. 477-491.
- [4] Chevalier, J.A., and D. Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews", *Journal of Marketing Research*, 2006 (43:3), pp. 345-354.
- [5] Clemons, E.K., G.G. Gao, and L.M. Hitt, "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry", *Journal of Management Information Systems*, 2006 (23:2), pp. 149-171.

- [6] Darby, M.R., and E. Karni, "Free Competition and the Optimal Amount of Fraud", *The Journal of Law and Economics*, 1973 (16:1), pp. 67-88.
- [7] Dellarocas, C., and R. Narayan, "What Motivates Consumers to Review a Product Online? A Study of the Product-specific Antecedents of Online Movie Reviews", in *Workshop on Information Systems and Economics*, Evanston, IL, 2006.
- [8] Federman, A.D., and S. Keyhani, "Physicians' Participation in the Physicians' Quality Reporting Initiative and Their Perceptions of Its Impact on Quality of Care", *Health Policy*, 2011 (102:2), pp. 229-234.
- [9] Forman, C., A. Ghose, and B. Wiesenfeld, "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets", *Information Systems Research*, 2008 (19:3), pp. 291-313.
- [10] Gao, G.G., J.S. McCullough, R. Agarwal, and A.K. Jha, "A Changing Landscape of Physician Quality Reporting: Analysis of Patients' Online Ratings of Their Physicians over a 5-Year Period", *Journal of Medical Internet Research*, 2012 (14:1), e38.
- [11] Gao, G.G., B.N. Greenwood, R. Agarwal, and J.S. McCullough, "Vocal Minority and Silent Majority: How do Online Ratings Reflect Population Perceptions of Quality?" *MIS Quarterly*, 2015 (39:3), pp. 565-589.
- [12] Godes, D., and J.C. Silva, "Sequential and Temporal Dynamics of Online Opinion", *Marketing Science*, 2012 (31:3), pp. 448-473.
- [13] Gray, B.M., J.L. Vandergrift, G.G. Gao, J.S. McCullough, and R.S. Lipner, "Website Ratings of Physicians and Their Quality of Care", *JAMA Internal Medicine*, 2015 (175:2), pp. 291-293.
- [14] Gu, B., J. Park, and P. Konana, "Research Note—The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products", *Information Systems Research*, 2012 (23:1), pp. 182-196.
- [15] Hanauer, D.A., K. Zheng, D.C. Singer, A. Gebremariam, and M.M. Davis, "Public Awareness, Perception, and Use of Online Physician Rating Sites", *The Journal of the American Medical Association*, 2014 (311:7), pp. 734-735.
- [16] Irani, J.S., J.L. Middleton, R. Marfatia, E.T. Omana, and F. D'amico, "The Use of Electronic Health Records in the Exam Room and Patient Satisfaction: A Systematic Review", *The Journal of the American Board of Family Medicine*, 2009 (22:5), pp. 553-562.
- [17] Li, X., and L.M. Hitt, "Self-selection and Information Role of Online Product Reviews," *Information Systems Research*, 2008 (19:4), pp. 456-474.
- [18] Lu, S.F., and H. Rui, "Can We Trust Online Physician Ratings? Evidence from Cardiac Surgeons in Florida", *Management Science*, 2017 (64:6), pp. 2557-2573.
- [19] Mayzlin, D., Y. Dover, and J. Chevalier, "Promotional Reviews: An Empirical Investigation of Online Review Manipulation", *American Economic Review*, 2014 (104:8), pp. 2421-2455.
- [20] Mark, B.A., C.B. Jones, and L. Lindley, "An Examination Technical Efficiency, Quality, and Patient Safety in Acute Care Nursing Units", *Policy, Politics, & Nursing Practice*, 2009 (10:3), pp. 180-186.
- [21] Mudambi, S.M., and D. Schuff, "Research note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com", *MIS Quarterly*, 2010 (34:1), pp. 185-200.
- [22] Muchnik, L., S. Aral, and S.J. Taylor, "Social Influence Bias: A Randomized Experiment", *Science*, 2013 (341:6146), pp. 647-651.
- [23] Nielsen, F.Å., "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs" in *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts'*, 2011, pp. 93-98.
- [24] Ostroff, C., and N. Schmitt, "Configurations of Organizational Effectiveness and Efficiency," *Academic of Management Journal*, 1993 (36:6), pp. 1345-1361.
- [25] Saiffee, D.H., Z. Zheng, I.R. Bardhan, and A. Lahiri, "Are Online Reviews of Physicians Reliable Indicators of Clinical Outcomes?", *Workshop on Information Systems and Economics*, Seoul, South Korea, 2017.